

1                   **SUPPLEMENTARY DATA**  
2

3                   **Chromosome folding and prophage activation reveal specific genomic architecture for**  
4                   **intestinal bacteria**

5  
6                   **Authors**

7                   Quentin LAMY-BESNIER<sup>1,2</sup>, Amaury BIGNAUD<sup>2,3</sup>, Julian R. GARNEAU<sup>4</sup>, Marie  
8                   TITECAT<sup>1</sup>, Devon CONTI<sup>1,2,3</sup>, Alexandra VON STREMPEL<sup>5</sup>, Marc MONOT<sup>4</sup>, Bärbel  
9                   STECHER<sup>5,6</sup>, Romain KOSZUL<sup>2</sup>, Laurent DEBARBIEUX<sup>1</sup>, Martial MARBOUTY<sup>2</sup>

10  
11  
12                   **Supplementary Figure 1a and 1b:** Contact map of each bacterial strain of the OMM<sup>12</sup>  
13 consortium obtained from *in vitro* cultures.

14                   **Supplementary Figure 2:** Re-assembly of *B. animalis*, *F. plautii*, and *B. caecimuris*.

15                   **Supplementary Figure 3:** Signal of the secondary diagonals for the different bacteria of the  
16                   OMM<sup>12</sup> consortium.

17                   **Supplementary Figure 4:** Comparison of the contact maps (*in vitro* vs. *in vivo*) for the six  
18                   most abundant bacteria.

19                   **Supplementary Figure 5:** Hierarchical clustering of the different Hi-C replicates for the  
20                   different bacteria of the OMM<sup>12</sup> consortium using the software HiCrep.

21                   **Supplementary Figure 6a and 6b:** Contact maps of functional prophage candidates (+/- 50  
22                   kb).

23                   **Supplementary Figure 7:** Krona representation of the Kaiju annotation of the reads not  
24                   mapping on the OMM<sup>12</sup> strains' genomes.

25                   **Supplementary Figure 8:** Viral clustering of the 13 induced phages using vContact2.

26  
27                   **Supplementary Table 1:** Genomic libraries generated.

28                   **Supplementary Table 2:** Genbank accession numbers of the OMM12 bacteria genomes.

29                   **Supplementary Table 3:** Metrics of the assemblies obtained with virome reads that did not  
30                   map on the OMM12 strains.

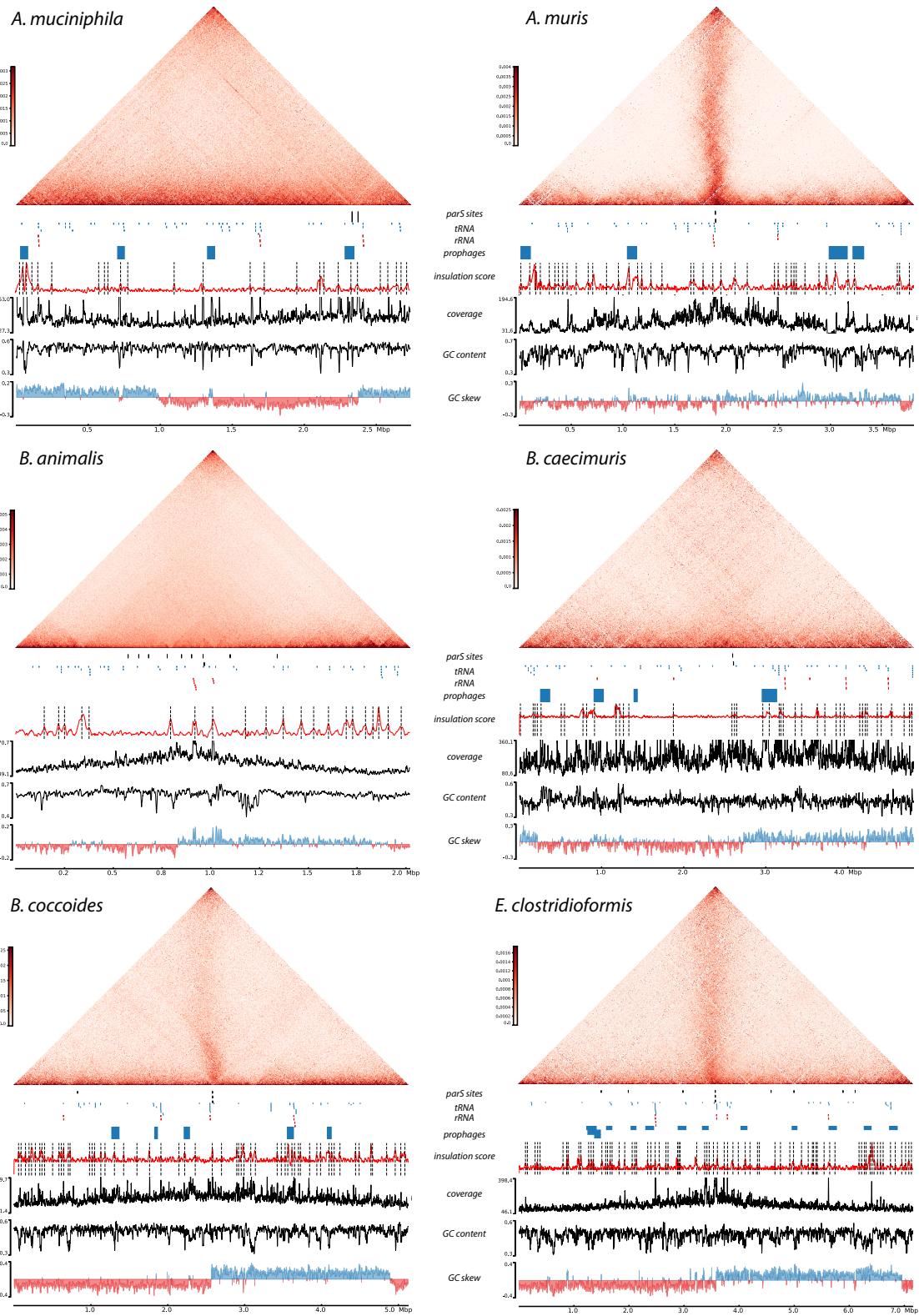
31                   **Supplementary Table 4:** Blast results of the contigs obtained by assembling non-mapping  
32                   reads.

33  
34  
35

36 **Supplementary Figure 1a and 1b: Contact map of each bacterial strain of the OMM<sup>12</sup>**  
 37 **consortium obtained from *in vitro* cultures.**

38 Each contact map is represented with associated genomic information: localization of *parS* sites  
 39 (green), tRNA (blue) and rRNA (red) on top with below prophage annotation (blue), coverage,  
 40 GC content and GC skew.

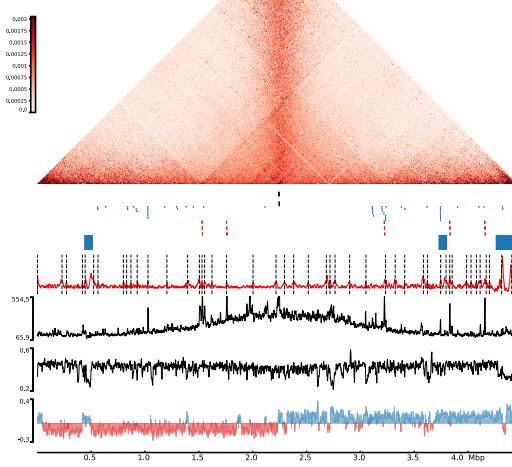
41



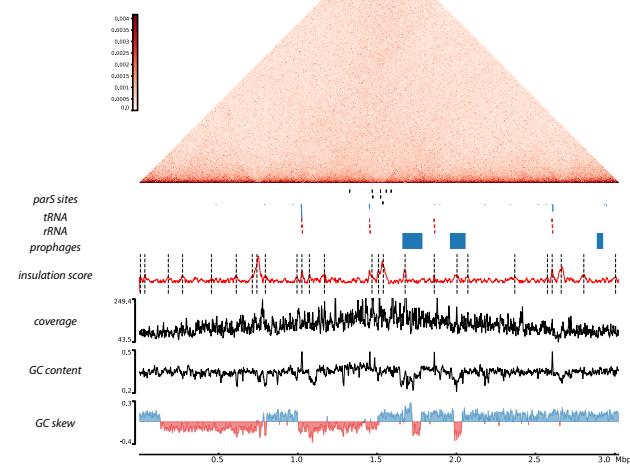
42

43

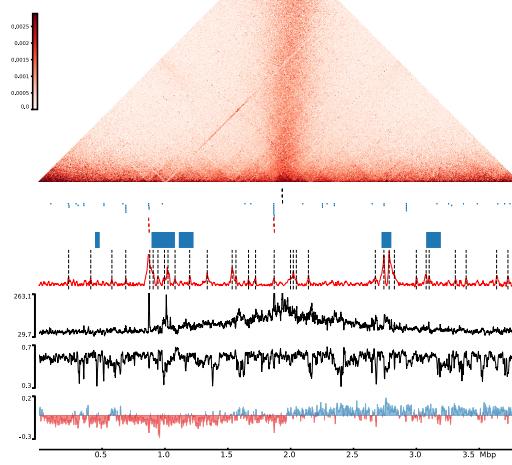
*C. innocuum*



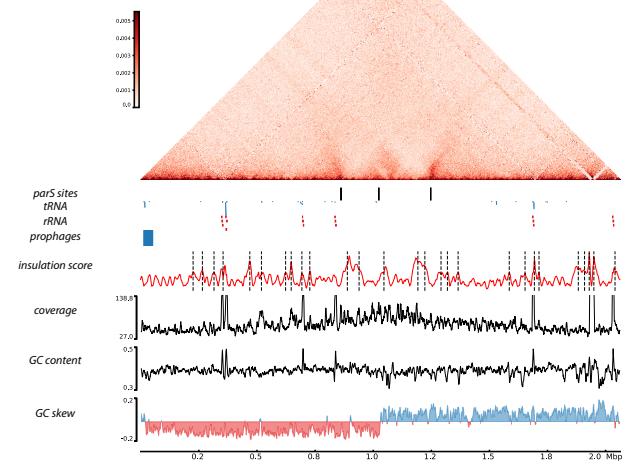
*E. faecalis*



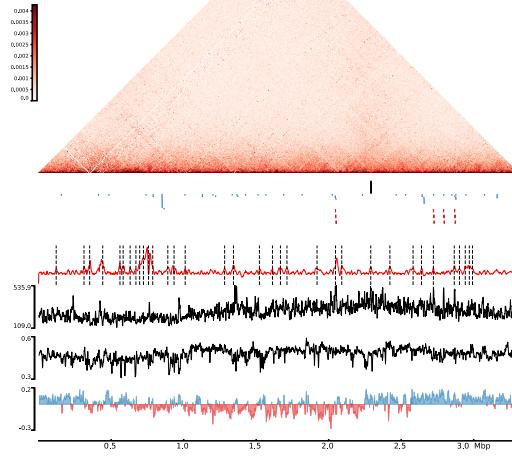
*F. plautii*



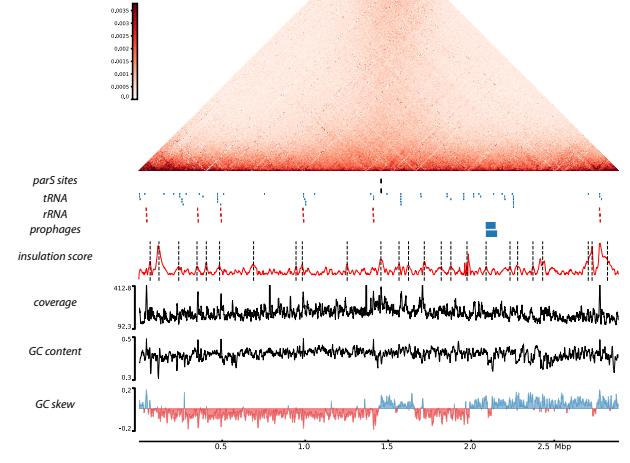
*L. reuteri*



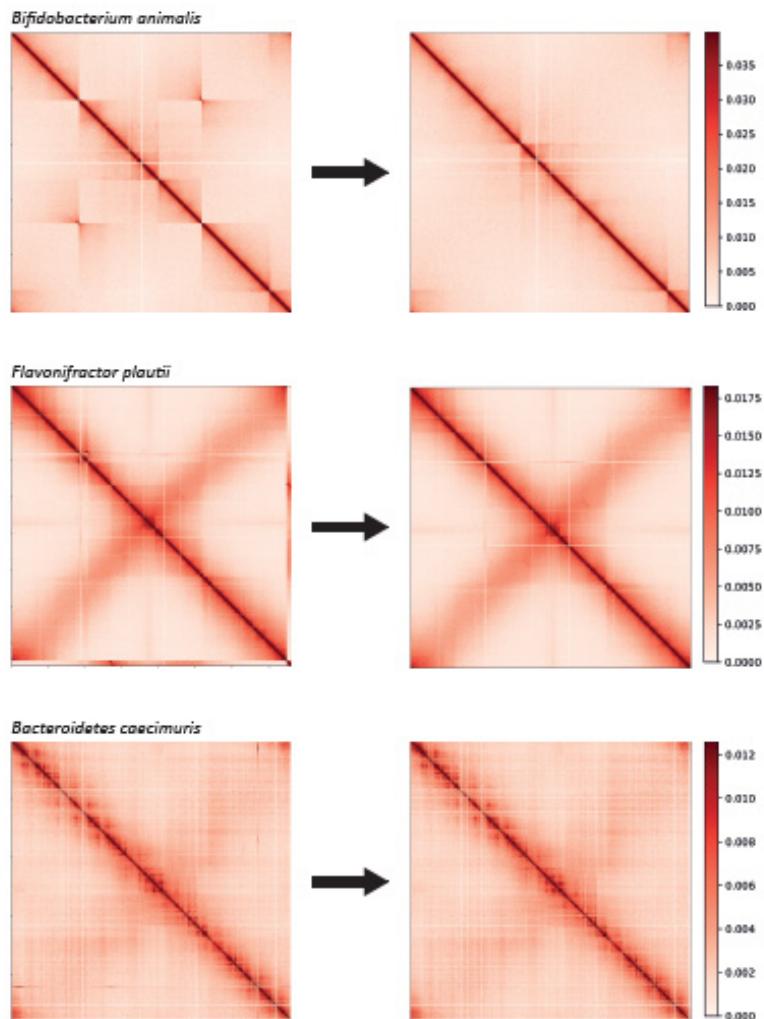
*M. intestinalis*



*T. muris*



45 **Supplementary Figure 2: Re-assembly of *B. animalis*, *F. plautii*, and *B. caecimuris*.**  
46 The contact maps are shown before (left) and after (right) Hi-C-based re-assembly.  
47

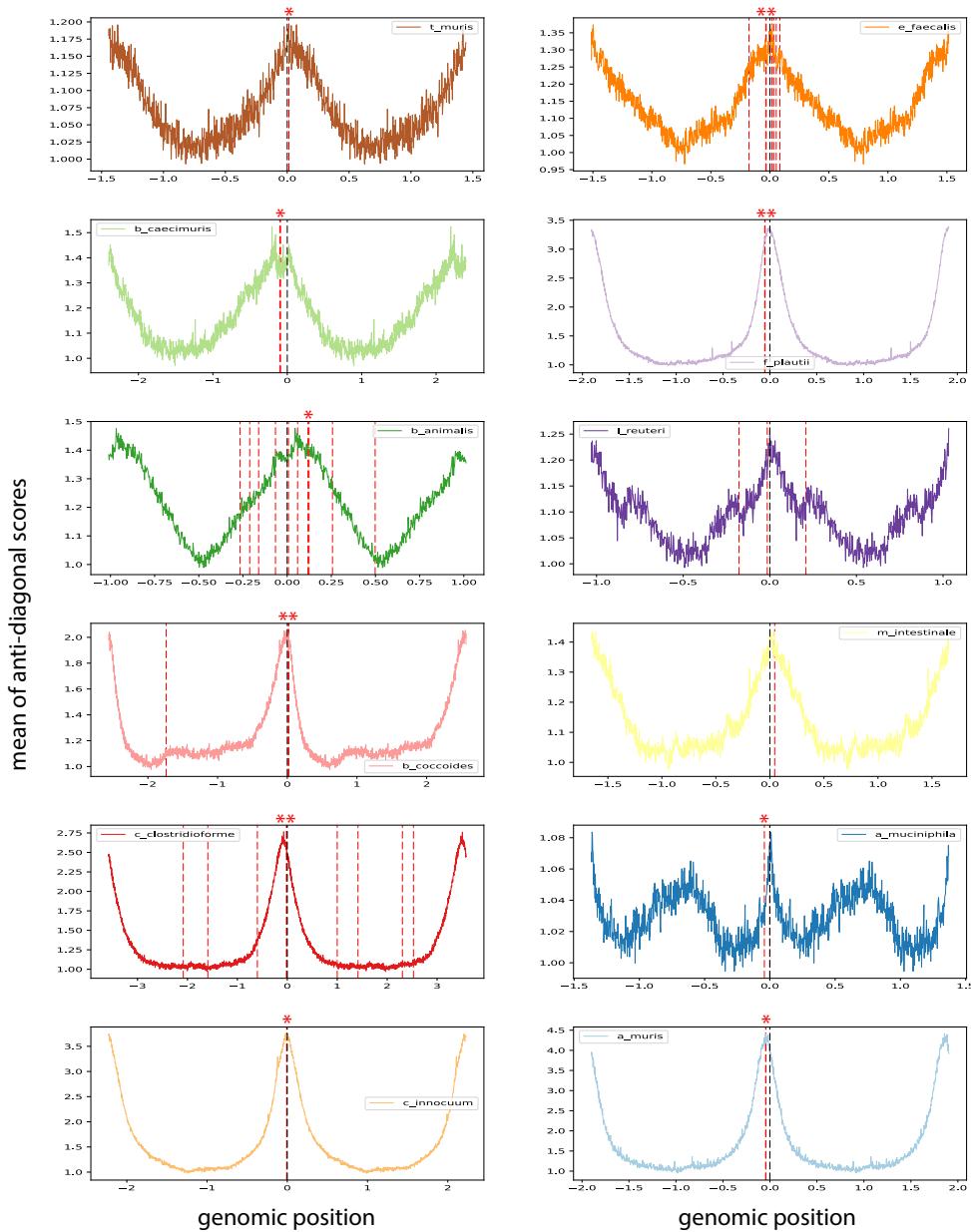


48  
49

50 **Supplementary Figure 3: Signal of the secondary diagonals for the twelve bacteria of the**  
 51 **OMM<sup>12</sup> consortium grown *in vitro*.**

52 Plots are centered on *ori*. Localisation of *parS* sites are indicated as red dashed lines. Red stars  
 53 indicate the presence of several *parS* sites in the same 5 kb window.

54

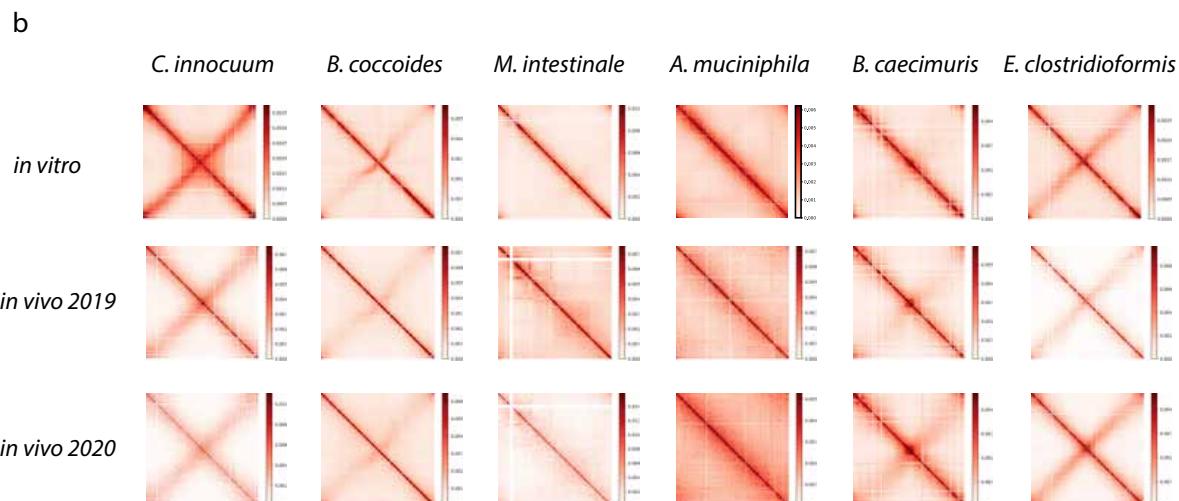
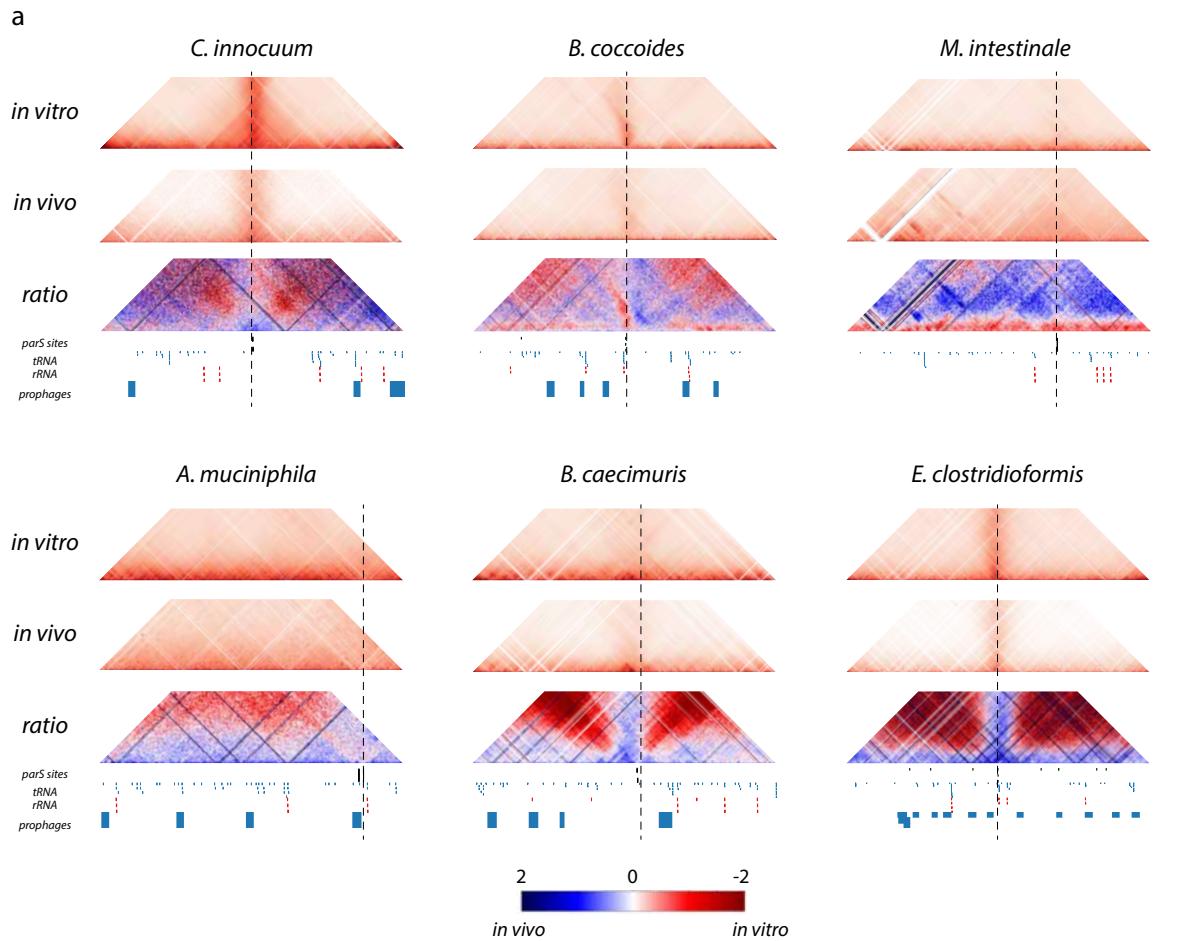


55  
 56

57 **Supplementary Figure 4: Comparison of the contact maps (*in vitro* vs. *in vivo*) for the six  
58 most abundant intestinal bacteria in OMM<sup>12</sup> mice.**

59 **a.** *In vitro* (5 kb bin), *in vivo* (5 kb bin) and ratio (Log2; 10 kb bin) of contact maps (*in vitro* vs.  
60 *in vivo*) obtained for the six most abundant bacteria in OMM<sup>12</sup> mice. Specific annotations are  
61 indicated under matrices. Origin of replication are indicated by dashed black lines. **b.** The three  
62 matrices (*in vitro* (up), *in vivo* 2019 (middle) and *in vivo* 2020 (bottom)) are shown on top of  
63 each other.

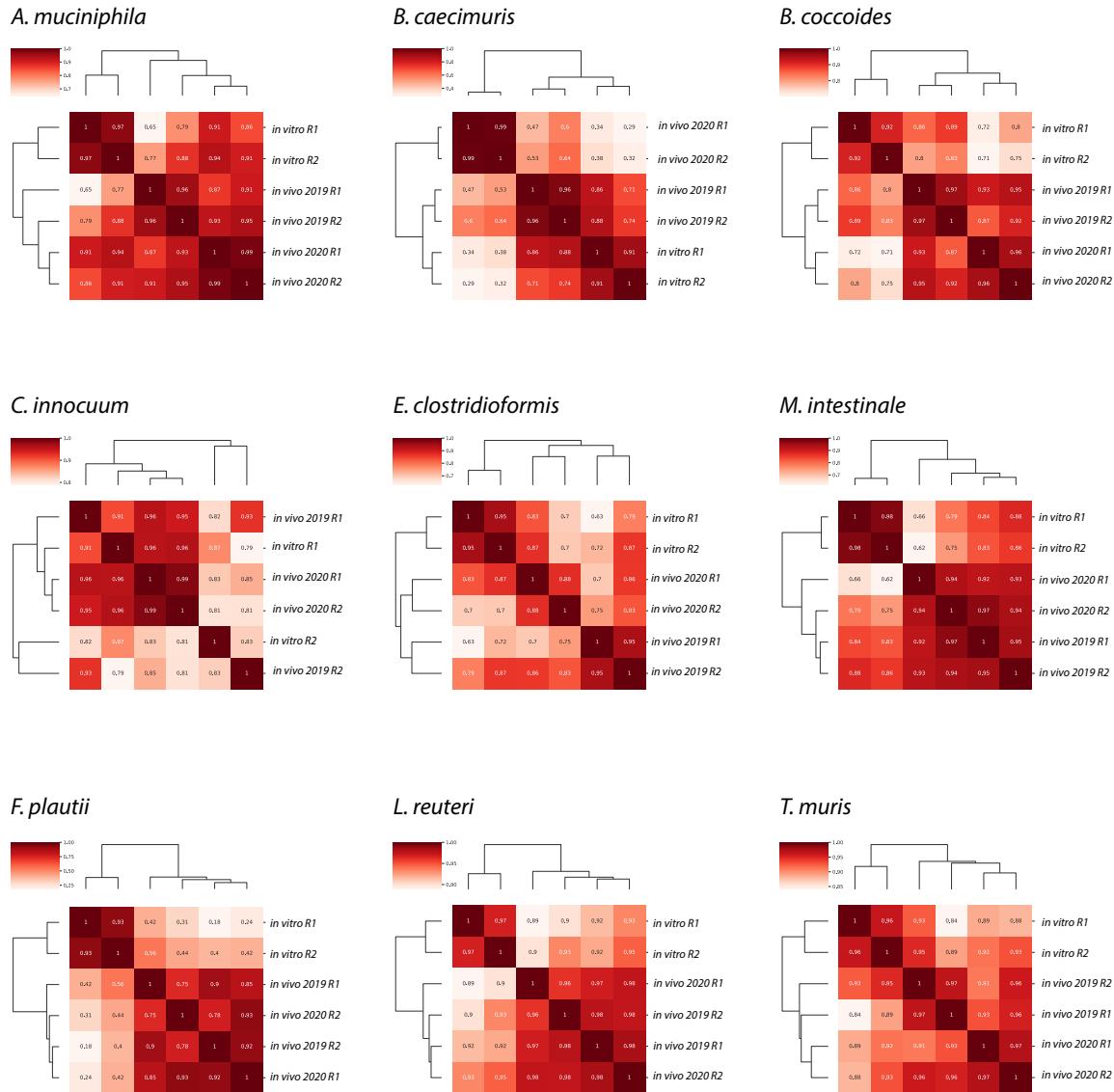
64



65

66

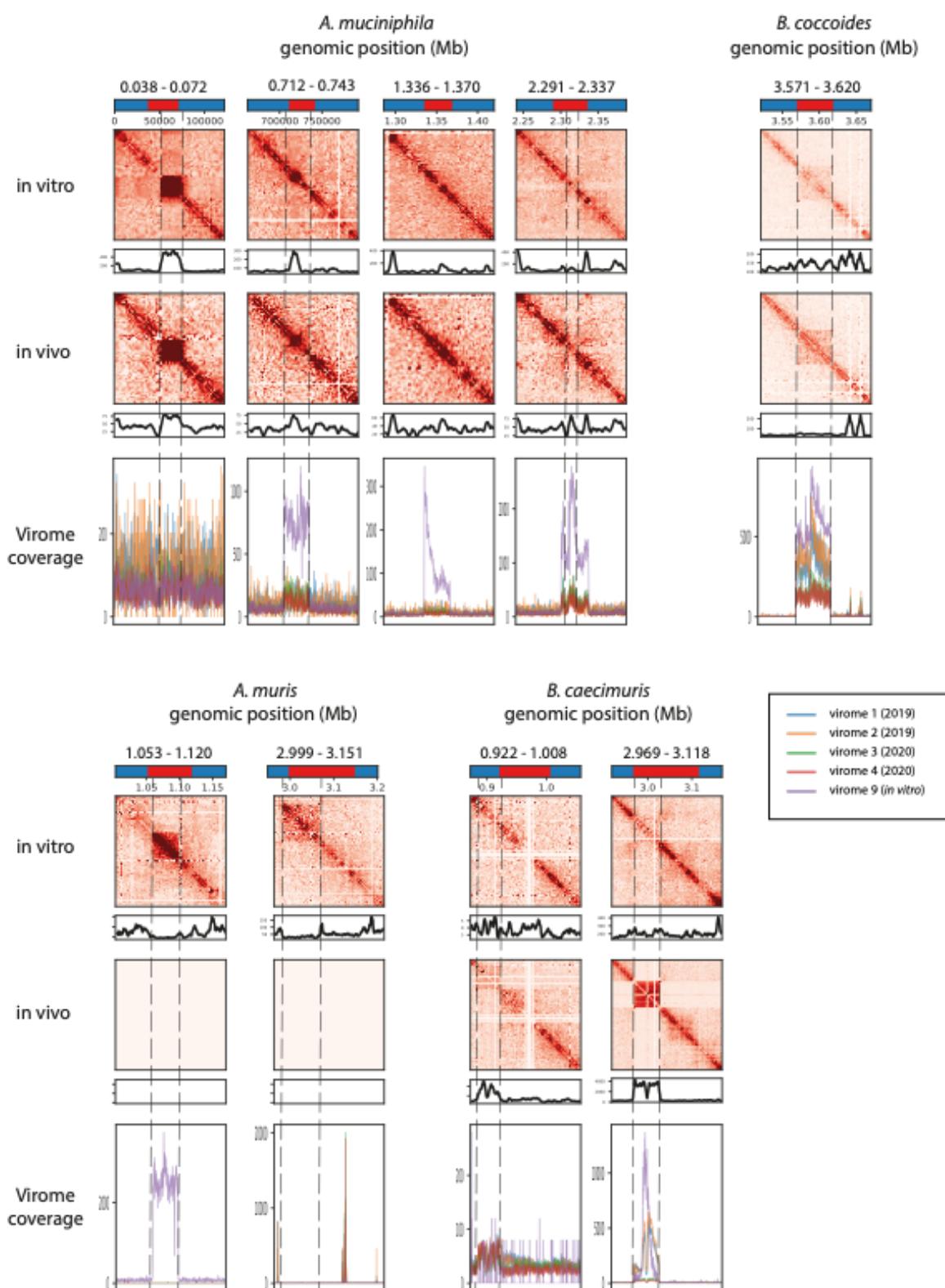
67      **Supplementary Figure 5: Hierarchical clustering of the different Hi-C replicates for the**  
 68      **different bacteria of the OMM<sup>12</sup> consortium using the software HiCrep.**  
 69

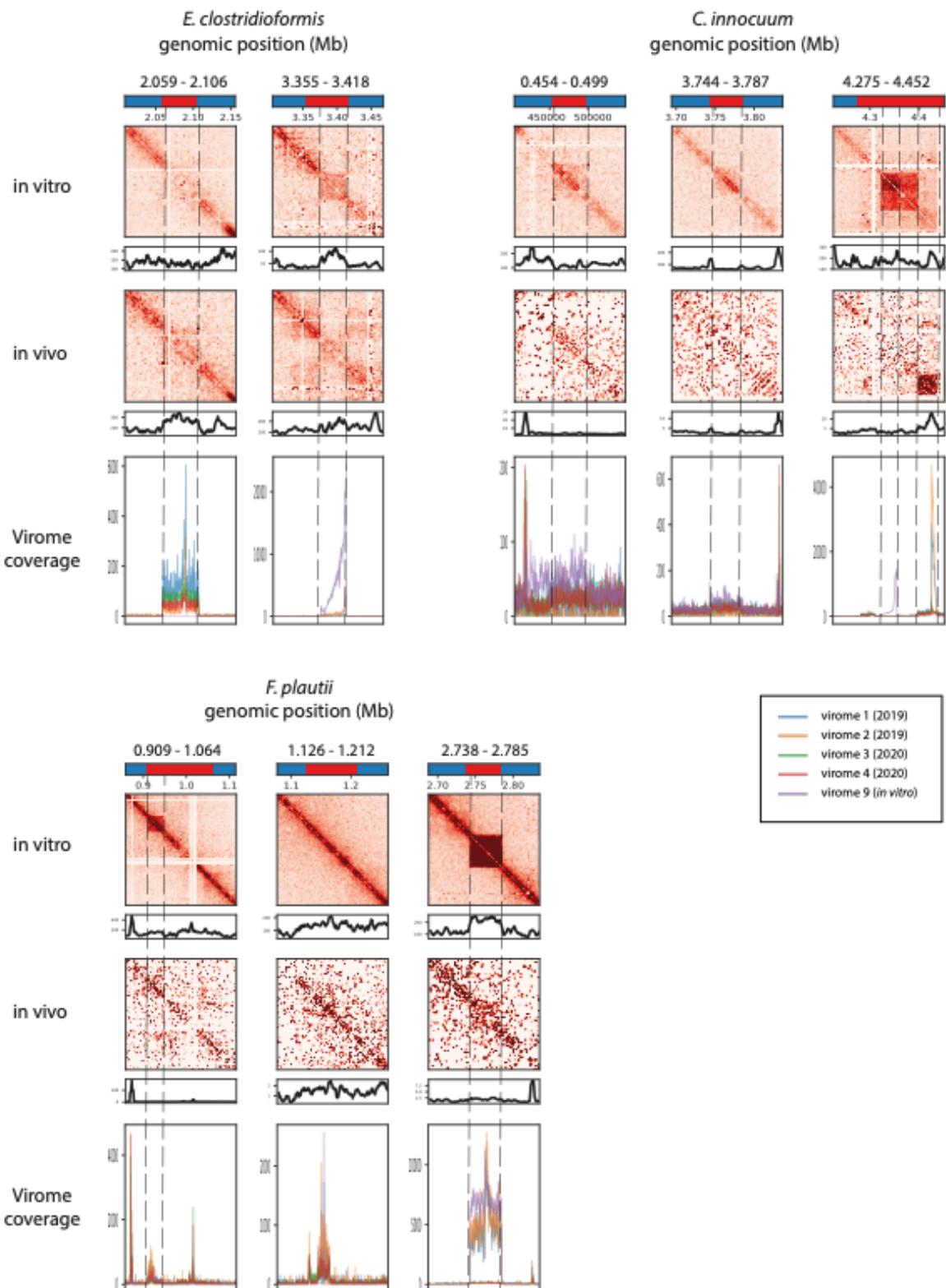


70  
71

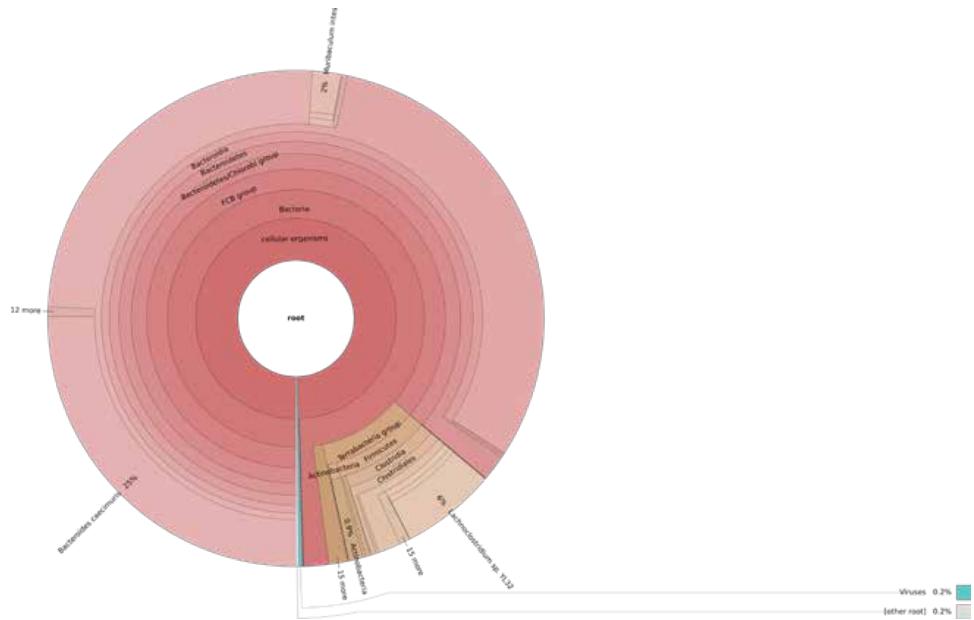
72 **Supplementary Figure 6a and 6b: Contact maps of functional prophage candidates (+/-  
73 50 kb).**

74 Genomic coordinates are indicated above contact maps while Hi-C coverage as well as virome  
75 data are indicated under. Dashed lines indicate Hi-C refinement of prophage coordinates. Blue  
76 and red regions indicate, respectively, chromosome and predicted prophages.  
77



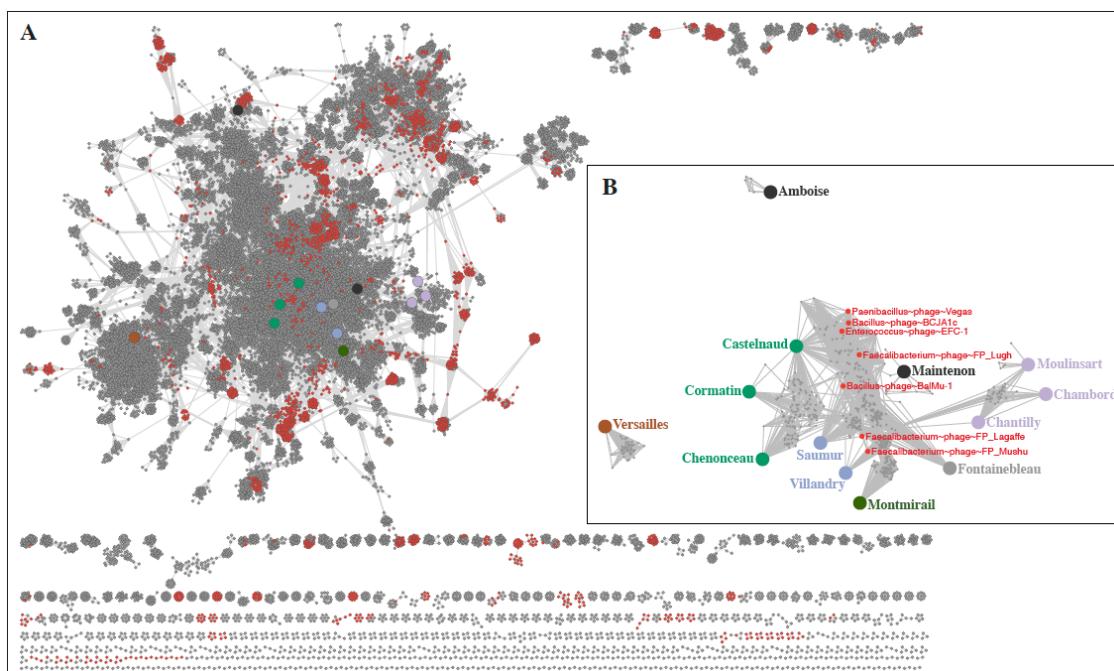


81      **Supplementary Figure 7: Krona representation of the Kaiju annotation of the reads not**  
82      **mapping on the OMM<sup>12</sup> strains' genomes.**  
83



84  
85

86 **Supplementary Figure 8: Viral clustering of the 13 induced prophages using vConctact2.**  
87 **a.** Viral cluster analysis with vConTACT2 using a gene-sharing network. The analysis was  
88 performed using genomes from the ViralRefSeq V.201 (red nodes) and CHVD (grey nodes)  
89 reference databases. Nodes for OMM<sup>12</sup> prophages were colored according to their respective  
90 host. **b.** Close-up view of the 13 OMM<sup>12</sup> inducible prophages and their direct (first level)  
91 neighbors in the network.  
92



93  
94

95      **Supplementary Table 1: Genomic libraries generated.**  
 96      a: libraries Virome\_5 and Virome\_6 were prepared with Accel-NGSTM 1S Plus DNA Library  
 97      Kit, and thus contain accurate data for both ssDNA and dsDNA viruses. b: spike 1 was  
 98      composed of 10<sup>8</sup> PFU/mL of phages CLB\_P1, CLB\_P2, CLB\_P3 and M13. c: spike 2 was  
 99      composed of 10<sup>7</sup> PFU/mL of phages CLB\_P1, CLB\_P2, CLB\_P3 and M13.  
 100

type	lib id	geno me	species	conditio ns	cage	sex	sampli ng date	raw reads (paired-end)	Accession number
Hi-C <i>in vitro</i>	OMM5	I46	<i>C.innocuum</i>	Lab	NA	NA	march -21	41 561 382	SRX15328653
	OMM5_r ep	I46	<i>C.innocuum</i>	Lab	NA	NA	July- 21	10 688 910	SRX15328654
	OMM6	YL58	<i>B.coccooides</i>	Lab	NA	NA	march -21	21 485 845	SRX15328665
	OMM6_r ep	YL58	<i>B.coccooides</i>	Lab	NA	NA	July21	2 257 624	SRX15328676
	OMM7	I48	<i>B.caecimuris</i>	Lab	NA	NA	march -21	93 139 653	SRX15328684
	OMM7_r ep	I48	<i>B.caecimuris</i>	Lab	NA	NA	July- 21	61 639 162	SRX15328685
	OMM8	YL27	<i>M.intestinal e</i>	Lab	NA	NA	march -21	31 275 362	SRX15328686
	OMM8_r ep	YL27	<i>M.intestinal e</i>	Lab	NA	NA	July- 21	9 298 553	SRX15328687
	OMM9	YL44	<i>A.muciniphil a</i>	Lab	NA	NA	march -21	40 820 347	SRX15328688
	OMM9_r ep	YL44	<i>A.muciniphil a</i>	Lab	NA	NA	July- 21	7 062 251	SRX15328689
	OMM10	YL32	<i>C.clostridiof orme</i>	Lab	NA	NA	march -21	62 518 175	SRX15328655
	OMM10_r ep	YL32	<i>C.clostridiof orme</i>	Lab	NA	NA	July- 21	6 408 840	SRX15328656
	OMM11	KB1	<i>E.faecalis</i>	Lab	NA	NA	march -21	12 250 166	SRX15328657
	OMM11_r ep	KB1	<i>E.faecalis</i>	Lab	NA	NA	July- 21	11 076 026	SRX15328658
	OMM12	YL31	<i>F.plautii</i>	Lab	NA	NA	march -21	10 544 208	SRX15328659
	OMM12_r ep	YL31	<i>F.plautii</i>	Lab	NA	NA	July- 21	17 539 144	SRX15328660
	OMM13	I49	<i>L.reuteri</i>	Lab	NA	NA	march -21	9 363 639	SRX15328661
	OMM13_r ep	I49	<i>L.reuteri</i>	Lab	NA	NA	July- 21	4 516 936	SRX15328662
	OMM14	YL2	<i>B.animalis</i>	Lab	NA	NA	march -21	7 526 155	SRX15328663
	OMM14_r ep	YL2	<i>B.animalis</i>	Lab	NA	NA	July- 21	4 669 716	SRX15328664
	OMM15	YL45	<i>T.muris</i>	Lab	NA	NA	march -21	8 827 292	SRX15328666
	OMM15_r ep	YL45	<i>T.muris</i>	Lab	NA	NA	July- 21	18 422 688	SRX15328667
	OMM16	KB18	<i>A.muris</i>	Lab	NA	NA	march -21	7 757 787	SRX15328668
	OMM16_r ep	KB18	<i>A.muris</i>	Lab	NA	NA	July- 21	9 033 311	SRX15328669
Hi-C <i>in vivo</i>	OMM1	mix	microbiota	microbi ota	3	male	sept- 19	101 182 905	SRX15328670
	OMM2	mix	microbiota	microbi ota	4	fema le	sept- 19	115 105 053	SRX15328671
	OMM3	mix	microbiota	microbi ota	1	male	may- 20	72 638 739	SRX15328672
	OMM4	mix	microbiota	microbi ota	2	fema le	may- 20	94 694 874	SRX15328673
Viro me	Virome1	mix	dsDNA	microbi ota	4	fema le	sept- 19	6 443 793 (2x35)	SRX15328674
	Virome2	mix	dsDNA	microbi ota	3	male	sept- 19	4 957 507 (2x35)	SRX15328675

	Virome3	mix	dsDNA	microbiota	1	male	may-20	108 807 445 (2x35 & 2x150)	SRX15328677
	Virome4	mix	dsDNA	microbiota	2	male	may-20	119 010 486 (2x35 & 2x150)	SRX15328678
	Virome5	mix	ssDNA <sup>a</sup> + spike 1 <sup>b</sup>	microbiota	breeding	mix	march -21	1 056 335 (2x150)	SRX15328679
	Virome6	mix	ssDNA <sup>a</sup> + spike 1 <sup>b</sup>	microbiota	breeding	mix	march -21	2 763 300 (2x150)	SRX15328680
	Virome7	mix	dsDNA + spike 2 <sup>c</sup>	microbiota	breeding	mix	march -21	5 174 852 (2x150)	SRX15328681
	Virome8	mix	dsDNA + spike 2 <sup>c</sup>	microbiota	breeding	mix	march -21	4 339 389 (2x150)	SRX15328682
	Virome9	mix (in vitro)	dsDNA	microbiota	breeding	NA	oct-21	6 979 488 (2x150)	SRX15328683

101

102

103      **Supplementary Table 2: Genbank accession numbers of the OMM12 bacteria genomes.**  
 104

Genus	Species	Strain	GenBank assembly accession
<b>akkermansia</b>	<b>muciniphila</b>	<b>YL44</b>	<b>GCA_016697425.1</b>
<b>acutalibacter</b>	<b>muris</b>	<b>KB18</b>	<b>GCA_016697365.1</b>
<b>bifidobacterium</b>	<b>animalis</b>	<b>YL2</b>	<b>GCA_023278655.1</b>
<b>bacteroides</b>	<b>caecimuris</b>	<b>I48</b>	<b>GCA_023277905.1</b>
<b>blautia</b>	<b>pseudococcoides</b>	<b>YL58</b>	<b>GCA_016696745.1</b>
<b>clostridium</b>	<b>innocuum</b>	<b>I46</b>	<b>GCA_016697325.1</b>
<b>enterocloster</b>	<b>clostridoformis</b>	<b>YL32</b>	<b>GCA_016696785.1</b>
<b>enterococcus</b>	<b>faecalis</b>	<b>KB1</b>	<b>GCA_016696825.1</b>
<b>flavonifractor</b>	<b>plautii</b>	<b>YL31</b>	<b>GCA_023277885.1</b>
<b>limosilactobacillus</b>	<b>reuteri</b>	<b>I49</b>	<b>GCA_016697045.1</b>
<b>muribaculum</b>	<b>intestinale</b>	<b>YL27</b>	<b>GCA_016696845.1</b>
<b>turicimonas</b>	<b>muris</b>	<b>YL45</b>	<b>GCA_016696765.1</b>

105  
 106  
 107      **Supplementary Table 3: Metrics of the assemblies obtained with virome reads that did**  
 108      **not map on the OMM<sup>12</sup> strains.**  
 109      Both SPAdes and Megahit were used.  
 110

Sample name	Sample info	Number of non mapping reads	SPAdes		Megahit	
			Number of contigs	Number of contigs >5kb	Number of contigs	Number of contigs >5kb
#1	2019, 2x35	358 904	68	2	59	3
#2	2019, 2x35	467 489	46	2	50	5
#3-35	2020, 2x35	6 530 360	10766	1	8214	0
#3-150	2020, 2x150	791 870	1611	2	2075	0
#4-35	2020, 2x35	10 045 524	10241	0	5903	0
#4-150	2020, 2x150	471 135	3119	0	4320	0
merge	all above samples	18 665 282	14012	0	8721	0
merge #3	2020, 2x35+2x150	7 322 230	15107	0	5887	0
merge #4	2020, 2x35+2x150	10 516 659	11067	0	5903	0

112  
113  
114  
115

**Supplementary Table 4: Blast results of the contigs obtained by assembling non-mapping reads.**

software	Contig name	Associated sample	Size	Best blast hit	Blast %coverage	Blast %id	Blast e-value
SPADES	CS1	#1 (2019)	24 646	B_caecimuris	100	99.90	0
	CS2	#1 (2019)	22 127	B_caecimuris	100	99.78	0
	CS3	#2 (2019)	53 027	B_caecimuris	100	99.8	0
	CS4	#2 (2019)	7 265	C.innocuum	100	100	0
	CS6	#3-150 (2020, 150)	6 639	B_caecimuris	100	99.95	0
	CS7	#3-150 (2020, 150)	6 369	B_caecimuris	100	99.76	0
	CS8	#4-150 (2020, 150)	5 854	B_caecimuris	100	99.79	0
	CM1	#1 (2019)	33 237	B.caecimuris	100	99.84	0
MegaHit	CM2	#1 (2019)	9 945	B.caecimuris	100	99.82	0
	CM3	#1 (2019)	6 574	B.caecimuris	99	99.71	0
	CM4	#2 (2019)	20 229	B.caecimuris	100	99.83	0
	CM5	#2 (2019)	9 690	B.caecimuris	100	99.96	0
	CM6	#2 (2019)	8 089	B.caecimuris	100	99.84	0
	CM7	#2 (2019)	6 839	B.caecimuris	99	99.9	0
	CM8	#2 (2019)	5 999	C.innocuum	100	100	0

116